Hyperspherical von Mises-Fisher Mixture (HvMF) Modelling of High Angular Resolution Diffusion MRI*

Abhir Bhalerao[1] and Carl-Fredrik Westin[2]

[1] Department of Computer Science, University of Warwick, Coventry CV4 7AL
abhir.bhalerao@dcs.warwick.ac.uk
[2] Laboratory of Mathematics in Imaging, Harvard Medical School, Brigham and Womens Hospital, Boston MA 02115
westin@bwh.harvard.edu

Abstract. A mapping of unit vectors onto a 5D hypersphere is used to model and partition ODFs from HARDI data. This mapping has a number of useful and interesting properties and we make a link to interpretation of the second order spherical harmonic decompositions of HARDI data. The paper presents the working theory and experiments of using a von Mises-Fisher mixture model for directional samples. The MLE of the second moment of the HvMF pdf can also be related to fractional anisotropy. We perform error analysis of the estimation scheme in single and multi-fibre regions and then show how a penalised-likelihood model selection method can be employed to differentiate single and multiple fibre regions.

1 Introduction

The directional dependence of diffusion of water molecules in brain white matter is the basis of DWI and a widely adopted non-invasive method for elucidating white matter fibre directions and, through tractography, inferring connectivity between brain regions [1]. DWI involves the acquisition of a set of images, in a small number of directions, and reconstructing the Gaussian diffusion by estimating the diffusion tensor. For regions containing a bundle of fibres all oriented in the same direction, the diffusion tensor model can characterise local *apparent* diffusion with as few as 6 directions. In the regions where the fibres bifurcate, cross or are adjacent to white-matter surfaces, the single tensor model is insufficient. High angular resolution diffusion imaging (HARDI) [2] can detect more precisely the variation of diffusion along different directions. For a given (larger) set of gradient directions, HARDI imaging can be analysed to produce samples of a pdf of diffusion over the surface of a sphere – the radial marginal of the pdf of the particle displacements. However, characterisation of the local geometry

* This work is partly funded by NIH Grants R01MH074794 and P41RR13218. We are also grateful to Ulas Ziyan at CSAIL MIT, and Gordon Kindlmann (LMI) for their help and insights into this work.

N. Ayache, S. Ourselin, A. Maeder (Eds.): MICCAI 2007, Part I, LNCS 4791, pp. 236–243, 2007.

given such measurements, called the orientation distribution function or ODF, is much less clear than in diffusion tensor imaging.

Frank [3] proposed the use of spherical harmonics (SH) to characterize the local geometry of the diffusivity. A notable finding of his was that single fibre regions show up in the 2nd order harmonics set: Y_2^{-2}, Y_2^{-1}, Y_2^0, and Y_2^1, Y_2^2, whilst the order 4 functions can add further information about multiple fibre regions. He made proposals for the separation of these regions according to the prominence of a particular channel. In this work, we show the relationship between the projection of ODF samples on to the same basis set and that a particular linear combinations of the second order projections are just components of the rank tensor mapping. Thus a way to determine the principal diffusion direction (PDD) when using SHs to model and reconstruct the HARDI data is revealed.

Since the ODF is a distribution function, a natural way to model it is by pdfs. However, commonly used Gaussian models do not extend to the sphere in a straightforward way because of the problem of "wrapping" of 3D angles modulo π and 2π. McGraw [4] used the von Mises-Fisher (vMF) distribution to parameterise the ODF. To capture the structure of multiple fibre voxels, they fitted a mixture of vMF density functions with pairs of antipodal modes, with directions $\{\boldsymbol{\mu}_1, -\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, -\boldsymbol{\mu}_2\}$, and went on to give expressions for scalar metrics of the parameterization (entropy), and distance metrics between pairs of mixtures using Riemannian Exp and Log maps for the purposes of interpolation. We build on this work by considering only unimodal and bimodal mixtures through a mapping of samples drawn from the ODF to a 5D representation which is free from the ambiguities associated with sign flips of vectors direction in 3D. This alleviates the need for introducing pseudo-modes into the fit. Such a representation of orientation was originally proposed by Knutsson [5] and has been used for filtering and optical flow analysis in vision. Recently Rieger and van Vliet [6] presented new insights into such orientation spaces and their properties. We show that these properties are important to measurements in diffusion imaging.

2 Theory

2.1 Hyperspherical von Mises-Fisher Distributions (HvMF)

The von Mises-Fisher (vMF) is the analog of the Gaussian distribution on a sphere and is parameterised by a principal direction (the mean direction $\boldsymbol{\mu}$) and a *concentration* parameter, κ. These distributions extends to arbitrary dimensions, p, though rarely are hyperspherical vMFs considered:

$$g_p(\boldsymbol{x}|\boldsymbol{\mu}, k) = c(\kappa)e^{\kappa \boldsymbol{\mu}^T \boldsymbol{x}}, \qquad c(\kappa) = \frac{\kappa^{p/2-1}}{((2\pi)^{p/2} I_{p/2-1}(\kappa))}, \tag{1}$$

where the normalisation factor, $c(\kappa)$, contains a Bessel function of the first kind to a fractional order. g_p is bell-shaped with the general form $e^{b\cos(\psi)}$, where the exponent will have the range $[-b, b]$ and ψ is the angle difference between the *direction* of the sample $\boldsymbol{x}$ and the mean direction $\boldsymbol{\mu}$. A set of vectors which point

more or less in the same direction would have a vMF pdf which is unimodal and symmetric around the mean direction. The relationship between the spread, b, and the variance or second-moment of the samples is less straightforward and non-linear. However, larger b values *concentrate* the probability mass around the mean direction.

2.2 A Double-Angle Representation for 3D Orientation

In directional statistics, antipodal vectors are regarded as the same. This ambiguity is elegantly removed in 2D by *angle doubling* or, in general, by taking outer products to form 2nd order tensors i.e. $\boldsymbol{x} \rightarrow \boldsymbol{x}\boldsymbol{x}^T$ where $\boldsymbol{x} = (x_1, x_2, x_3) \in \mathbb{R}^3$. The dimensionality of this space can be reduced by restricting the trace of this tensor to be 1 to produce the 5D mapping [5,6]. In spherical polar coordinates,

$$M_5(r, \theta, \phi) \rightarrow r(s, t, u, v, w), \tag{2}$$
$$s = \sin^2\theta \cos 2\phi, \qquad t = \sin^2\theta \sin 2\phi,$$
$$u = \sin 2\theta \cos\phi, \qquad v = \sin 2\theta \sin\phi, \qquad w = \sqrt{3}(\cos^2\theta - \frac{1}{3}).$$

Although not explained in [6], we cannot robustly solve for (θ, ϕ) given any two coefficients in M_5. To accurately invert the mapping therefore, we have to reconstitute the implied tensor, $\boldsymbol{x}\boldsymbol{x}^T$, and calculate the direction of its principal eigen vector:

$$M_5^{-1} : r(s, t, u, v, w) \rightarrow r(\theta, \phi), \qquad \boldsymbol{x}\boldsymbol{x}^T = \begin{pmatrix} s + x_2^2 & \frac{t}{2} & \frac{u}{2} \\ \frac{t}{2} & x_3^2 - \frac{s+w\sqrt{3}}{2} & \frac{v}{2} \\ \frac{u}{2} & \frac{v}{2} & \frac{w\sqrt{3}+1}{3} \end{pmatrix}. \tag{3}$$

We note also that the above 5D mapping is equivalent to weighted amounts of selected 2nd order spherical harmonic basis functions:

$$(s, t, u, v, w) = (\frac{1}{3}Y_2^2, 8Y_2^{-2}, \frac{2}{3}Y_2^1, 4Y_2^{-1}, \frac{2}{\sqrt{3}}Y_2^0). \tag{4}$$

This is used below to estimate PDD directly from projections of HARDI samples.

2.3 Maximum Likelihood Estimates of HvMF Parameters

HvMF pdfs are parameterised by two parameters: the mean $\boldsymbol{\mu}$ and the concentration parameter κ. These can be used to model the apparent diffusion of homogeneous fibre regions.

Given a set of independent sample vectors, $\boldsymbol{x}_i, i = 1..n$, believed to be from $g_p(\boldsymbol{x}|\boldsymbol{\mu}, \kappa)$, the maximum likelihood estimate of the mean is obtained by the sum of the vectors divided by the length of the sum. It can be shown that the MLE of the concentration parameter, κ, is then obtained as follows:

$$\boldsymbol{r} = \sum_i^n \boldsymbol{x}_i, \qquad \hat{\boldsymbol{\mu}} = \frac{\boldsymbol{r}}{\| \boldsymbol{r} \|}, \qquad \hat{\kappa} = A^{-1}(\bar{R}) \approx \frac{\bar{R}p - \bar{R}^3}{1 - \bar{R}^2}, \qquad \bar{R} = \frac{1}{n}\hat{\boldsymbol{\mu}}^T\boldsymbol{r} \tag{5}$$

where $A(\kappa) = I_{p/2}(\kappa)/I_{p/2-1}(\kappa)$, is the ratio of modified Bessel functions of the first kind to fractional orders (see Sra [7]).

2.4 Mixture Modelling: Fitting by EM and Model Selection

A HvMF mixture allows the modelling of more than one principal direction but an algorithm such as Expectation Maximization (EM) is needed to perform the parameter estimation. If we now assume that a set of samples, $\boldsymbol{x}_i, i = 1...n$, are now drawn from a m-mode mixture distribution,

$$G_p(\boldsymbol{x}|w, \Theta) = \sum_j^m w_j g_p(\boldsymbol{x}|\Theta_j) \tag{6}$$

with a convex set of weights, $\sum_j^m w_j = 1$ and each mode is parameterised by $\Theta_j = \{\boldsymbol{\mu}_j, \kappa_j\}$, then it can be shown that MLE for the mixture are given by the update equations (abbreviating notation for brevity):

$$\hat{\mu}_j^{t+1} = \frac{\sum_i^n P_{ij}^t \hat{\kappa}_{ij}^t \boldsymbol{x}_i}{\| \sum_i^n P_{ij}^t \hat{\kappa}_j^t \boldsymbol{x}_i \|} \qquad A(\hat{\kappa}_j^{t+1}) = \frac{\sum_i^n P_{ij}^t \boldsymbol{x}_i^T \boldsymbol{\mu}_j^t}{\sum_i^n P_{ij}^t} \qquad \hat{w}_j^{t+1} = \frac{1}{n} \sum_i^n P_{ij}^t \tag{7}$$

The posterior value at step t, P_{ij}^t, is calculated in the usual way for an EM algorithm from the expectation of the data, $\boldsymbol{x}_i$ given the current weights and parameter estimates, w_j, Θ^t.

Given a MLE fit to the samples, a parsimonious way to determine what number of modes m is best, is to use a model selection criterion such as the Akaike information criterion (AIC). The AIC is a number based on the log-likelihood of the data penalised by the number of parameters used to model the distribution. Thus, for the HvMF, for m modes, we have $m(p + 2)$ parameters altogether (remembering the mean is p-dimensional). To select the model, we minimize for m

$$AIC(m) = -2 \sum_i^n \log G_p(x_i|\hat{w}, \hat{\Theta}) + 2m(p + 2). \tag{8}$$

For our purposes, only $AIC(1)$ and $AIC(2)$ need be compared to select between single or multi-fibre regions.

2.5 Relationships Between Variance of ODF and a Measure of Anisotropy

The ML estimate of variance of the transformed set of samples, $\boldsymbol{X}_i \in \mathbb{S}^5$, can be used to characterise the anisotropy of the a Rank 1 tensor estimate. The ML estimate of variance in $\mathbb{S}^5$ is given by $1 - \bar{R}$ (from equation 5). $\bar{R}$ is the variation along the mean $\boldsymbol{\mu}$, while the spread is the perpendicular projection of the vector of length $n\bar{R}$ along $\hat{\boldsymbol{\mu}}$ (figure 1):

$$\boldsymbol{r} = n\bar{R}\boldsymbol{\mu}, \qquad var(\boldsymbol{X}) = \| \boldsymbol{I} - \boldsymbol{r}\boldsymbol{r}^T \| . \tag{9}$$

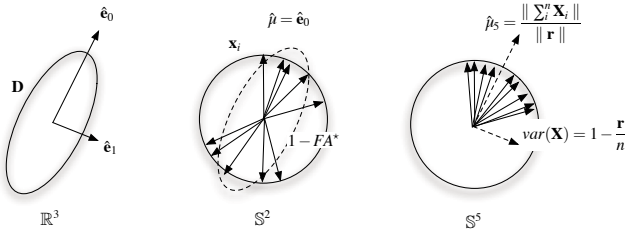


Fig. 1. This figure relates the common tensor diffusion model (left) with the mapping from the ODF (middle) and the 5D orientation mapping (right). Notice we have the ambiguity resulting from vectors in $\mathbb{R}^3$ being used to describe diffusion. The reason this is problematic is that diffusion requires a tensor quantity for its correct characterisation. The sign problem is resolved in the 5D space where all the vectors are concentrated around one direction.

We can define a modified fractional (FA) anisotropy related to standard FA:

$$FA = \sqrt{\frac{3}{2}} \frac{\| \boldsymbol{D} - \frac{trace\boldsymbol{D}}{3} \boldsymbol{I} \|}{\| \boldsymbol{D} \|}, \qquad FA^\star = \frac{\| \sum_i^n \boldsymbol{x}\boldsymbol{x}^T - \lambda_1 \boldsymbol{e}_1 \boldsymbol{e}_1^T \|}{\| \sum_i^n \boldsymbol{x}\boldsymbol{x}^T \|}. \tag{10}$$

where $\| \, . \, \|$ is the tensor norm. The geometric interpretation of FA* is that in FA the average diffusion in all directions is removed from the tensor to make it traceless, whereas only the diffusion in the principal direction, $\lambda_1 \boldsymbol{e}_i \boldsymbol{e}_1^T$ is taken away in FA*. So, FA* will be larger than FA when $\lambda_1 \gg \lambda_2, \lambda_3$ but smaller when $\lambda_1 \approx \lambda_2, \lambda_3$[1].

3 Experiments and Discussion

We synthesized noisy ODFs from apparent diffusion of a tensor model with $S_0 = 1$ and $b = 700$ s/mm^2 with the approximate q-ball reconstruction technique outlined in [8]. These ODFs were then randomly sampled to give $\boldsymbol{x}_i, i = 1..1024$. Then having used M_5, we performed 10 iterations of EM according to the ML update steps outlined above. In the illustrative results in figure 2 and the error analyses (figure 2), we used the same number of modes (m) as the number of synthesizing tensors used (ie. $m = 1$ or $m = 2$). For moderate SNR ratios, e.g. SNR $= 64$, the fitting is robust. Error analysis for $m = 1$ indicates that even in low SNRs the average estimation stays below 10^o. For the two tensor case, we plotted the minimum angle error between either principal axis and either HvMF modal direction (4 possible correspondences are tried). The average of the minimum and maximum errors, which is an upper bound on this error, was then plotted.

We used HARDI data containing 120 gradient directions with $b = 700$s/mm^2 for further experiments. The ML estimates of the M_5 mean (of the HvMF eqn. 5) was used to create a PDD map (by decomposing the reconstituted tensor eqn. 3).

[1] $\star$ as it symbolises the spread of the vectors.

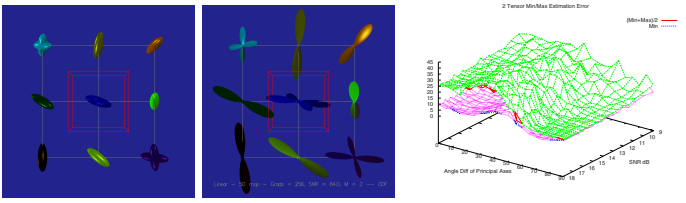
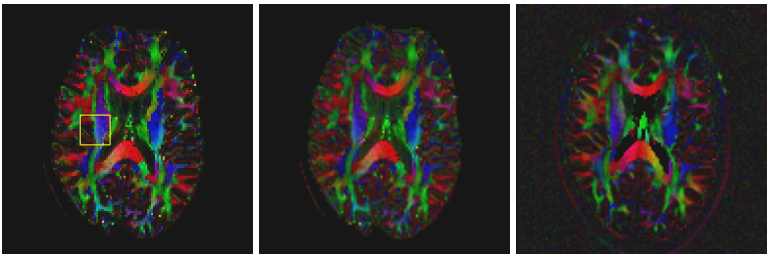


Fig. 2. HvMF fitting to samples drawn from noisy single and two tensor diffusion models (1024 samples used for EM estimation with SNR = 64dB). Tensor eccentricity is fixed at $(1, \frac{1}{3}, \frac{1}{3})$. Left figure depicts the single and muti-tensor voxels. Centre figure shows vHMF pdfs as isosurface: $\boldsymbol{x}G(\boldsymbol{x}|\Theta)$. Right figure, angle errors plot between two closest matching directions.



(a) (b) (c)

Fig. 3. Comparison of FA maps and PDDs obtained in different ways: (a) Standard FA with PDD shown in colours from a tensor fit to data (roi region used used below); (b) FA* produced using rank 1 spread measure (see equation 10); (c) FA from reconstituted tensor given 2nd order spherical harmonic basis expansion. Data contains 120 HARDI measurements (b is 700 s/mm^2). Note the similarities of all three maps which are obtained in different ways.

This map was weighted by the FA* and compared with coloured PDD and FA obtained by a standard least-squares tensor fitting (figures 3(a) and (b)). The resultant maps are indistinguishable other than, as expected, FA* being slightly greater in isotropic regions. Figure 3(c) shows estimates of PDD and FA obtained from HARDI data by the identity in equation 4. The HARDI measurements were interpolated using a cosine weighting kernel and then integrated with the 2nd order SH basis set, Y_2^m, by Monte Carlo integration. The coefficient images were then combined and a tensor reconstituted using the inverse M_5 mapping to yield the PDD and FA. The results are identical to FA of a standard tensor fit.

The images in figure 4 show 3D visualisations of a region from 1 slice of HARDI data showing the HvMF model selected fits to each voxel in comparison with single tensor fitting near the ventricle boundary. Qualitatively the results appear to be satisfactory but it is hard to judge whether the model selection is sensible. The HvMF is detecting planar and isotropic diffusion but by generally

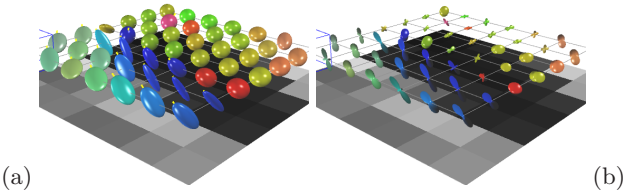


Fig. 4. Model selection by AIC on HvMF mixtures: (a) Single tensor estimates in small region on white-matter/CSF boundary; (b) HvMF estimates in same boundary region as in (a)

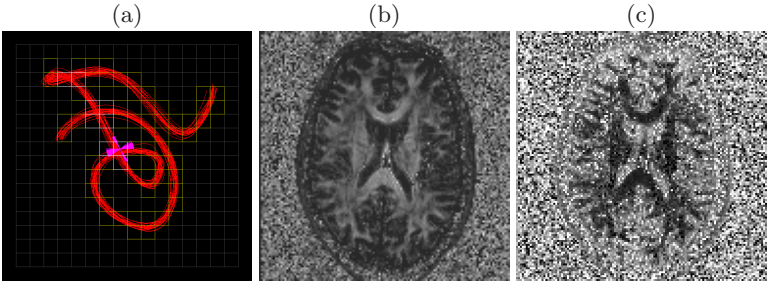


Fig. 5. (a) 2D synthetic example using vMF model selection (white=2) (yellow=1). (b) Log-likelihood of unimodal fit of H-vMF. (c) Log-likelihood of bimodal fit of H-vMF weighted by normalised squared difference in weights of two modal amplitudes (see text).

under fitting the data and preferring bimodal fits over a unimodal fit with small κ (note for $\kappa = 0$, the distribution is uniform). The AIC might be responsible as it less severe at penalising free parameters than say the Bayesian Information Criterion (BIC). It might be also that including a single parameter uniform component, $\frac{w_0}{4\pi}$, as part of the mixture to model will filter out the background low probabilities. 2D synthetic data was proceed in figure 5 where each region was labelled as either unimodal (yellow) or bimodal (white): all crossing regions are correctly labelled white. In figures 5(a) and (b) the voxel log-likelihoods after HvMF fitting are shown for the two cases. As expected, in (a), the fit is good in places where the fibre bundles generally lie and the map resembles a map of FA. In figure 5(b), we weighted the log-likelihood by the amplitudes of the normalised squared difference between the two principal modes: $(w_0 - w_1)^2/(w_0 + w_1)^2$ which will weight down those regions where their is a dominant mode. The results show that complementary regions to (a) are favoured by the bimodal fitting. Overall, the results indicate that some form of decision based selection may be necessary for better discrimination than achieved here, as reported recently by Peled et al. [9], if HvMF fitting was to be used in multi-tensor tractography. Such investigations are on-going.

4 Conclusions

We have described a probability model for high angular diffusion data. A 5D orientation mapping which resolves the inherent ambiguities in describing directional samples on a sphere is employed. This enables us to describe general ODFs in a natural and continuous way. In other words, we take advantage of the tensor mapping that respects that diffusion is bidirectional without having to resort to a Gaussian model. We also outlined the connection between spherical harmonic analysis of HARDI samples and our orientation space and how PDD calculations are equivalent. Our analyses indicate that this could be a fruitful approach for partitioning single and multi-fibre diffusion from HARDI data.

References

1. Weickert, J., Hagen, H.: Visualization and Processing of Tensor Fields. Springer, Heidelberg (2006)
2. Tuch, D.S., Reese, T.G., Wiegell, M.R., Makris, N., Belliveau, J.W., Weeden, V.J.: High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. Magnetic Resonance in Medicine 48(4), 577–582 (2002)
3. Frank, L.R.: Characterization of Anisotropy in High Angular Resolution Diffusion-Weighted MRI. Magnetic Resonance in Medicine 47, 1083–1099 (2002)
4. McGraw, T., Vemuri, B., Yezierski, B., Mareci, T.: Segmentation of High Angular Resolution Diffusion MRI Modeled as a Field of von Mises-Fisher Mixtures. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3953, pp. 461–475. Springer, Heidelberg (2006)
5. Knutsson, H.: Producing a continuous and distance preserving 5-d vector repesentation of 3-d orientation. In: Proceedings of IEEE Computer Society Workshop on Computer Architecture for Pattern Analysis and Image Database Management, pp. 18–20. IEEE Computer Society Press, Los Alamitos (1985)
6. Rieger, B., van Vliet, L.J.: Representing Orientation in n-Dimensional Spaces. In: Petkov, N., Westenberg, M.A. (eds.) CAIP 2003. LNCS, vol. 2756, pp. 17–24. Springer, Heidelberg (2003)
7. Sra, S., Dhillon, I.S.: Modeling Data using Directional Distributions. Technical report, Department of Computer Sciences, University of Texas (2002)
8. Bergman, O., Kindlmann, G., Lundervold, A., Westin, C.F.: Diffusion k-tensor Estimation from Q-ball Imaging Using Discretized Principal Axes. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4191, pp. 268–275. Springer, Heidelberg (2006)
9. Peled, S., Friman, O., Jolesz, F., Westin, C.-F.: Geometrically constrained two-tensor model for crossing tract in DWI. Magnetic Resonance Imaging 24, 1263–1270 (2006)